

#2

OIPE

RAW SEQUENCE LISTING

DATE: 04/26/2001

PATENT APPLICATION: US/09/782,672

TIME: 17:59:51

Input Set : N:\Crf3\RULE60\09782672.txt

Output Set: N:\CRF3\04262001\I782672.raw

SEQUENCE LISTING

ENTERED

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Iverson, Brent

6 Georgiou, George

7 Chen, Gang

8 Olsen, Mark J.

9 Daugherty, Patrick S.

11 (ii) TITLE OF INVENTION: Directed Evolution of Enzymes and
12 Antibodies

14 (iii) NUMBER OF SEQUENCES: 53

16 (iv) CORRESPONDENCE ADDRESS:

17 (A) ADDRESSEE: ARNOLD, WHITE AND DURKEE

18 (B) STREET: P.O. Box 4433

19 (C) CITY: Houston

20 (D) STATE: Texas

21 (E) COUNTRY: USA

22 (F) ZIP: 77210-4433

24 (v) COMPUTER READABLE FORM:

25 (A) MEDIUM TYPE: Floppy disk

26 (B) COMPUTER: IBM PC compatible

27 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

28 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

30 (vi) CURRENT APPLICATION DATA:

C--> 31 (A) APPLICATION NUMBER: US/09/782,672

C--> 32 (B) FILING DATE: 12-Feb-2001

33 (C) CLASSIFICATION:

35 (vii) PRIOR APPLICATION DATA:

36 (A) APPLICATION NUMBER: 08/847,063

37 (B) FILING DATE:

41 (viii) ATTORNEY/AGENT INFORMATION:

42 (A) NAME: Highlander, Steven L.

43 (B) REGISTRATION NUMBER: 37,642

44 (C) REFERENCE/DOCKET NUMBER: UTSB620

46 (ix) TELECOMMUNICATION INFORMATION:

47 (A) TELEPHONE: (512) 418-3000

48 (B) TELEFAX: (713) 789-2679

51 (2) INFORMATION FOR SEQ ID NO: 1:

53 (i) SEQUENCE CHARACTERISTICS:

54 (A) LENGTH: 780 base pairs

55 (B) TYPE: nucleic acid

56 (C) STRANDEDNESS: single

57 (D) TOPOLOGY: linear

60 (ix) FEATURE:

61 (A) NAME/KEY: CDS

62 (B) LOCATION: 1..780

64 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

66 GAA GTT CAA CTG CAA CAG TCT GGT CCT GAA TTG GTT AAA CCT GGC GCC

48

RAW SEQUENCE LISTING

DATE: 04/26/2001

PATENT APPLICATION: US/09/782,672

TIME: 17:59:51

Input Set : N:\Crf3\RULE60\09782672.txt

Output Set: N:\CRF3\04262001\I782672.raw

67	Glu	Val	Gln	Leu	Gln	Gln	Ser	Gly	Pro	Glu	Leu	Val	Lys	Pro	Gly	Ala	
68	1				5				10					15			
70	TCT	GTG	CGC	ATG	TCC	TGC	AAA	TCC	TCA	GGG	TAC	ATT	TTC	ACC	GAC	TTC	96
71	Ser	Val	Arg	Met	Ser	Cys	Lys	Ser	Ser	Gly	Tyr	Ile	Phe	Thr	Asp	Phe	
72				20				25					30				
74	TAC	ATG	AAT	TGG	GTT	CGC	CAG	TCT	CAT	GGT	AAG	TCT	CTA	GAC	TAC	ATC	144
75	Tyr	Met	Asn	Trp	Val	Arg	Gln	Ser	His	Gly	Lys	Ser	Leu	Asp	Tyr	Ile	
76		35					40					45					
78	GGG	TAC	ATT	TCC	CCA	TAC	TCT	GGG	GTT	ACC	GGC	TAC	AAC	CAG	AAG	TTT	192
79	Gly	Tyr	Ile	Ser	Pro	Tyr	Ser	Gly	Val	Thr	Gly	Tyr	Asn	Gln	Lys	Phe	
80		50					55				60						
82	AAA	GGT	AAG	GCC	ACC	CTT	ACT	GTC	GAC	AAA	TCT	TCC	TCA	ACT	GCT	TAC	240
83	Lys	Gly	Lys	Ala	Thr	Leu	Thr	Val	Asp	Lys	Ser	Ser	Ser	Thr	Ala	Tyr	
84	65				70				75					80			
86	ATG	GAG	CTG	CGT	TCT	TTG	ACC	TCT	GAG	GAC	TCC	GCG	GTA	TAC	TAT	TGC	288
87	Met	Glu	Leu	Arg	Ser	Leu	Thr	Ser	Glu	Asp	Ser	Ala	Val	Tyr	Tyr	Cys	
88				85				90				95					
90	GCC	GGC	TCC	TCT	GGT	AAC	AAA	TGG	GCC	ATG	GAT	TAT	TGG	GGT	CAT	GGT	336
91	Ala	Gly	Ser	Ser	Gly	Asn	Lys	Trp	Ala	Met	Asp	Tyr	Trp	Gly	His	Gly	
92				100				105				110					
94	GCT	AGC	GTT	ACT	GTG	AGC	TCT	GGT	GGC	GGT	GGC	TCG	GGC	GGT	GGT	GGG	384
95	Ala	Ser	Val	Thr	Val	Ser	Ser	Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	
96		115					120				125						
98	TCG	GGT	GGC	GGC	GGA	TCA	GAC	ATA	GTA	CTG	ACC	CAG	TCT	CCA	GCT	TCT	432
99	Ser	Gly	Gly	Gly	Gly	Ser	Asp	Ile	Val	Leu	Thr	Gln	Ser	Pro	Ala	Ser	
100		130				135					140						
102	TTG	GCT	GTG	TCT	CTA	GGA	CAA	AGG	GCC	ACG	ATA	TCC	TGC	CGA	TCC	AGC	480
103	Leu	Ala	Val	Ser	Leu	Gly	Gln	Arg	Ala	Thr	Ile	Ser	Cys	Arg	Ser	Ser	
104	145				150				155					160			
106	CAA	AGT	CTC	GTA	CAT	TCT	AAT	GGT	AAT	ACT	TAT	CTG	AAC	TGG	TAC	CAA	528
107	Gln	Ser	Leu	Val	His	Ser	Asn	Gly	Asn	Thr	Tyr	Leu	Asn	Trp	Tyr	Gln	
108				165				170				175					
110	CAG	AAA	CCA	GGA	CAG	CCA	CCC	AAG	CTT	CTC	ATC	TAT	AAG	GTA	TCC	AAC	576
111	Gln	Lys	Pro	Gly	Gln	Pro	Pro	Lys	Leu	Leu	Ile	Tyr	Lys	Val	Ser	Asn	
112			180					185				190					
114	CGA	TTC	TCT	GGA	GTC	CCT	GCC	AGG	TTC	AGT	GGC	AGT	GGG	TCT	GAG	TCA	624
115	Arg	Phe	Ser	Gly	Val	Pro	Ala	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Glu	Ser	
116		195					200				205						
118	GAC	TTC	ACC	CTC	ACC	ATC	GAT	CCT	GTG	GAG	GAA	GAT	GAT	GCT	GCA	ATA	672
119	Asp	Phe	Thr	Leu	Thr	Ile	Asp	Pro	Val	Glu	Glu	Asp	Asp	Ala	Ala	Ile	
120		210				215					220						
122	TAT	TAC	TGT	AGC	CAA	ACT	ACG	CAT	GTT	CCA	CCC	ACG	TTC	GGC	TCG	GGG	720
123	Tyr	Tyr	Cys	Ser	Gln	Thr	Thr	His	Val	Pro	Pro	Thr	Phe	Gly	Ser	Gly	
124	225				230				235					240			
126	ACC	AAG	CTG	GAG	CTG	AAA	CGT	GCT	AGC	CAG	CCA	GAA	CTC	GCC	CCG	GAA	768
127	Thr	Lys	Leu	Glu	Leu	Lys	Arg	Ala	Ser	Gln	Pro	Glu	Leu	Ala	Pro	Glu	
128			245					250				255					
130	GAC	CCC	GAG	GAC													780
131	Asp	Pro	Glu	Asp													

RAW SEQUENCE LISTING

DATE: 04/26/2001

PATENT APPLICATION: US/09/782,672

TIME: 17:59:51

Input Set : N:\Crf3\RULE60\09782672.txt

Output Set: N:\CRF3\04262001\I782672.raw

```

132          260
135 (2) INFORMATION FOR SEQ ID NO: 2:
137   (i) SEQUENCE CHARACTERISTICS:
138       (A) LENGTH: 260 amino acids
139       (B) TYPE: amino acid
140       (D) TOPOLOGY: linear
142   (ii) MOLECULE TYPE: protein
144   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
146 Glu Val Gln Leu Gln Gln Ser Gly Pro Glu Leu Val Lys Pro Gly Ala
147   1          5          10          15
149 Ser Val Arg Met Ser Cys Lys Ser Ser Gly Tyr Ile Phe Thr Asp Phe
150          20          25          30
152 Tyr Met Asn Trp Val Arg Gln Ser His Gly Lys Ser Leu Asp Tyr Ile
153          35          40          45
155 Gly Tyr Ile Ser Pro Tyr Ser Gly Val Thr Gly Tyr Asn Gln Lys Phe
156          50          55          60
158 Lys Gly Lys Ala Thr Leu Thr Val Asp Lys Ser Ser Ser Thr Ala Tyr
159          65          70          75          80
161 Met Glu Leu Arg Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys
162          85          90          95
164 Ala Gly Ser Ser Gly Asn Lys Trp Ala Met Asp Tyr Trp Gly His Gly
165          100         105         110
167 Ala Ser Val Thr Val Ser Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
168          115         120         125
170 Ser Gly Gly Gly Gly Ser Asp Ile Val Leu Thr Gln Ser Pro Ala Ser
171          130         135         140
173 Leu Ala Val Ser Leu Gly Gln Arg Ala Thr Ile Ser Cys Arg Ser Ser
174          145         150         155         160
176 Gln Ser Leu Val His Ser Asn Gly Asn Thr Tyr Leu Asn Trp Tyr Gln
177          165         170         175
179 Gln Lys Pro Gly Gln Pro Pro Lys Leu Leu Ile Tyr Lys Val Ser Asn
180          180         185         190
182 Arg Phe Ser Gly Val Pro Ala Arg Phe Ser Gly Ser Gly Ser Glu Ser
183          195         200         205
185 Asp Phe Thr Leu Thr Ile Asp Pro Val Glu Glu Asp Asp Ala Ala Ile
186          210         215         220
188 Tyr Tyr Cys Ser Gln Thr Thr His Val Pro Pro Thr Phe Gly Ser Gly
189          225         230         235         240
191 Thr Lys Leu Glu Leu Lys Arg Ala Ser Gln Pro Glu Leu Ala Pro Glu
192          245         250         255
194 Asp Pro Glu Asp
195          260
198 (2) INFORMATION FOR SEQ ID NO: 3:
200   (i) SEQUENCE CHARACTERISTICS:
201       (A) LENGTH: 18 base pairs
202       (B) TYPE: nucleic acid
203       (C) STRANDEDNESS: single
204       (D) TOPOLOGY: linear
206   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

RAW SEQUENCE LISTING DATE: 04/26/2001
 PATENT APPLICATION: US/09/782,672 TIME: 17:59:52

Input Set : N:\Crif3\RULE60\09782672.txt
 Output Set: N:\CRF3\04262001\I782672.raw

```

208 TGGACCAACA ACATCGGT                                     18
211 (2) INFORMATION FOR SEQ ID NO: 4:
213     (i) SEQUENCE CHARACTERISTICS:
214         (A) LENGTH: 26 base pairs
215         (B) TYPE: nucleic acid
216         (C) STRANDEDNESS: single
217         (D) TOPOLOGY: linear
219     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
221 CCCATATCAC CAGCTCACCG TCTTTC                             26
224 (2) INFORMATION FOR SEQ ID NO: 5:
226     (i) SEQUENCE CHARACTERISTICS:
227         (A) LENGTH: 32 base pairs
228         (B) TYPE: nucleic acid
229         (C) STRANDEDNESS: single
230         (D) TOPOLOGY: linear
232     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
234 GACCCCGAGG ACTAACGTCT TCGAATAAAT AC                     32
237 (2) INFORMATION FOR SEQ ID NO: 6:
239     (i) SEQUENCE CHARACTERISTICS:
240         (A) LENGTH: 25 base pairs
241         (B) TYPE: nucleic acid
242         (C) STRANDEDNESS: single
243         (D) TOPOLOGY: linear
245     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
247 CCGAATTCGT TTGAACATGC CTAAC                             25
250 (2) INFORMATION FOR SEQ ID NO: 7:
252     (i) SEQUENCE CHARACTERISTICS:
253         (A) LENGTH: 29 base pairs
254         (B) TYPE: nucleic acid
255         (C) STRANDEDNESS: single
256         (D) TOPOLOGY: linear
258     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
260 CGGAATTCGT GCGCAACACG ATGAAGCTC                         29
263 (2) INFORMATION FOR SEQ ID NO: 8:
265     (i) SEQUENCE CHARACTERISTICS:
266         (A) LENGTH: 31 base pairs
267         (B) TYPE: nucleic acid
268         (C) STRANDEDNESS: single
269         (D) TOPOLOGY: linear
271     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
273 AGGGCATGCA AGGGCACCAA TAACTGCCTT A                       31
276 (2) INFORMATION FOR SEQ ID NO: 9:
278     (i) SEQUENCE CHARACTERISTICS:
279         (A) LENGTH: 26 base pairs
280         (B) TYPE: nucleic acid
281         (C) STRANDEDNESS: single
282         (D) TOPOLOGY: linear
284     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
286 TTGGCTGCAG TAATATATTG CAGCAT                             26

```

RAW SEQUENCE LISTING DATE: 04/26/2001
 PATENT APPLICATION: US/09/782,672 TIME: 17:59:52

Input Set : N:\Crf3\RULE60\09782672.txt
 Output Set: N:\CRF3\04262001\I782672.raw

```

289 (2) INFORMATION FOR SEQ ID NO: 10:
291   (i) SEQUENCE CHARACTERISTICS:
292       (A) LENGTH: 31 base pairs
293       (B) TYPE: nucleic acid
294       (C) STRANDEDNESS: single
295       (D) TOPOLOGY: linear
297   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
299 TGCAATATAT TACTGCAGCC AAACACGCA T           31
302 (2) INFORMATION FOR SEQ ID NO: 11:
304   (i) SEQUENCE CHARACTERISTICS:
305       (A) LENGTH: 30 base pairs
306       (B) TYPE: nucleic acid
307       (C) STRANDEDNESS: single
308       (D) TOPOLOGY: linear
310   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
312 CGGCAGTTTC TGCAGATATA TTCGCAAGAT           30
315 (2) INFORMATION FOR SEQ ID NO: 12:
317   (i) SEQUENCE CHARACTERISTICS:
318       (A) LENGTH: 31 base pairs
319       (B) TYPE: nucleic acid
320       (C) STRANDEDNESS: single
321       (D) TOPOLOGY: linear
323   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
325 CTTGCGAATA TATCTGCAGA AACTGCCGGA A           31
328 (2) INFORMATION FOR SEQ ID NO: 13:
330   (i) SEQUENCE CHARACTERISTICS:
331       (A) LENGTH: 40 base pairs
332       (B) TYPE: nucleic acid
333       (C) STRANDEDNESS: single
334       (D) TOPOLOGY: linear
336   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:
338 ACGCCACATC TTGCGAATAT ATCTGCAGAA ACTGCCGGAA           40
341 (2) INFORMATION FOR SEQ ID NO: 14:
343   (i) SEQUENCE CHARACTERISTICS:
344       (A) LENGTH: 18 base pairs
345       (B) TYPE: nucleic acid
346       (C) STRANDEDNESS: single
347       (D) TOPOLOGY: linear
349   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:
351 CAGGGTACAT TTTCACCG                           18
354 (2) INFORMATION FOR SEQ ID NO: 15:
356   (i) SEQUENCE CHARACTERISTICS:
357       (A) LENGTH: 46 base pairs
358       (B) TYPE: nucleic acid
359       (C) STRANDEDNESS: single
360       (D) TOPOLOGY: linear
362   (ix) FEATURE:
363       (A) NAME/KEY: modified_base
364       (B) LOCATION: one-of(13, 14, 22, 23, 28, 29)

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/782,672

DATE: 04/26/2001

TIME: 17:59:53

Input Set : N:\Crf3\RULE60\09782672.txt

Output Set: N:\CRF3\04262001\I782672.raw

L:31 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:32 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]